

Supplementary Material for HaHeAE: Learning Generalisable Joint Representations of Human Hand and Head Movements in Extended Reality

Zhiming Hu, Guanhua Zhang, Zheming Yin, Daniel Häufle, Syn Schmitt, Andreas Bulling

A. Reconstruction Error Distributions

To analyse the distributions of different methods' reconstruction errors, we calculated the cumulative distribution function (CDF) of the hand and head reconstruction errors, respectively, for performance comparison. The higher the CDF curve, the better the reconstruction performance. We can see from Figure 1 that our method achieves better performance than the baseline methods in terms of both hand and head reconstruction error distributions on all the datasets (EgoBody, ADT, and GIMO). These results further demonstrate that our method outperforms other methods in reconstruction quality.

B. Clustering Performance of Different Methods

To compare the clustering performance of different methods, we first calculated the semantic representations of the hand-head movements on the test set of EgoBody using different methods and then identified hand-head clusters based on these representations, respectively. HDBSCAN was applied to perform clustering and cosine similarity was used as the distance metric. We employed two commonly used clustering metrics to evaluate the clustering performance of different methods:

- **Davies–Bouldin Index (DBI):** DBI evaluates clusters' compactness and separation and is calculated as the average similarity between each cluster and its closest neighbour. The lower the DBI, the better the clustering quality.
- **Calinski-Harabasz Index (CHI):** CHI measures clusters' compactness and separation and is defined as the ratio of the between-cluster separation to the within-cluster dispersion. The higher the CHI, the better the clustering performance.

Table I shows the clustering performance of different methods on the EgoBody dataset. We can see that our method achieves superior performance than the baseline methods in terms of clustering quality.

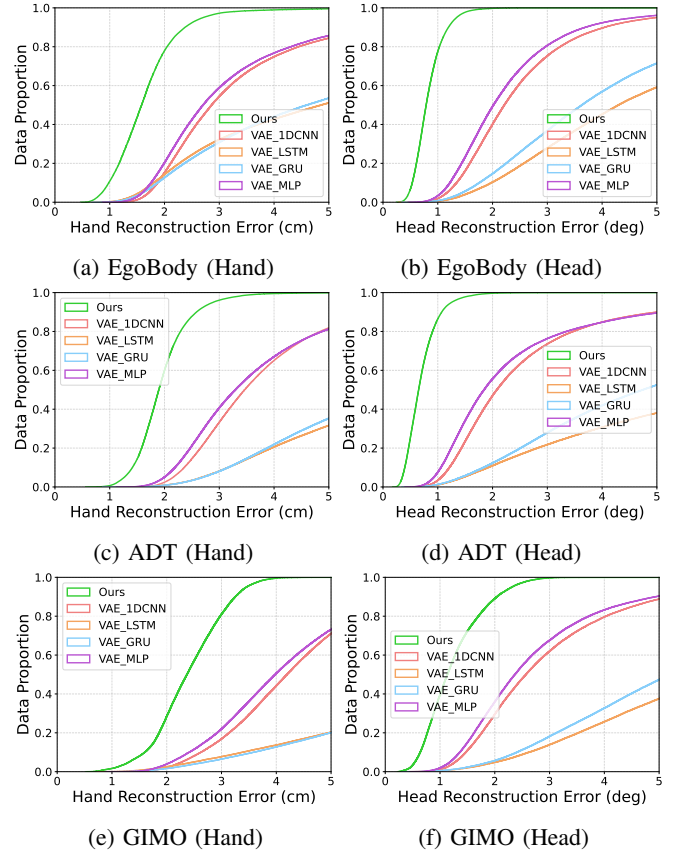


Fig. 1: The cumulative distribution functions of different methods' hand and head reconstruction errors on the EgoBody, ADT and GIMO datasets. Our method achieves better performance than other methods in terms of reconstruction error distributions.

TABLE I: The clustering performance of different methods on the EgoBody dataset. Our method outperforms other methods in terms of clustering quality.

	DBI ↓	CHI ↑
VAE_1DCNN	2.182	19.582
VAE_LSTM	1.343	26.626
VAE_GRU	1.367	28.892
VAE_MLP	1.837	29.485
Ours	1.141	37.970

Zhiming Hu, Guanhua Zhang, Zheming Yin, Syn Schmitt, and Andreas Bulling are with the University of Stuttgart, Germany. E-mail: {zhiming.hu@vis.uni-stuttgart.de, guanhua.zhang@vis.uni-stuttgart.de, st178328@stud.uni-stuttgart.de, schmitt@simtech.uni-stuttgart.de, andreas.bulling@vis.uni-stuttgart.de}. Daniel Häufle is with the University of Tübingen, Germany. E-mail: {daniel.haeufle@uni-tuebingen.de}. Daniel Häufle, Syn Schmitt, and Andreas Bulling are with the Center for Bionic Intelligence Tuebingen Stuttgart (BITS), Germany. Zhiming Hu is the corresponding author and is with the Hong Kong University of Science and Technology (Guangzhou), China.

Manuscript October 02, 2024.